



ASp

la revue du GERAS

7-10 | 1995

Actes du 16e colloque du GERAS

Le TGV Transmanche : création d'un outil de formation et de validation fondé sur la reconnaissance vocale

Pascale Fade



Édition électronique

URL : <http://journals.openedition.org/asp/3940>

DOI : 10.4000/asp.3940

ISSN : 2108-6354

Éditeur

Groupe d'étude et de recherche en anglais de spécialité

Édition imprimée

Date de publication : 1 décembre 1995

Pagination : 325-349

ISSN : 1246-8185

Référence électronique

Pascale Fade, « Le TGV Transmanche : création d'un outil de formation et de validation fondé sur la reconnaissance vocale », *ASp* [En ligne], 7-10 | 1995, mis en ligne le 12 novembre 2013, consulté le 19 avril 2019. URL : <http://journals.openedition.org/asp/3940> ; DOI : 10.4000/asp.3940

Ce document a été généré automatiquement le 19 avril 2019.

Tous droits réservés

Le TGV Transmanche : création d'un outil de formation et de validation fondé sur la reconnaissance vocale

Pascale Fade

- 1 Au congrès de la SAES à Strasbourg en mai 1992 et au colloque du GERAS à Grenoble en mars 1993, j'ai présenté le système de communication orale formalisée qui sera utilisé à bord du TGV Eurostar entre les conducteurs et le personnel au sol.
- 2 Dans cette intervention, après un bref rappel du contexte ferroviaire, je présenterai les choix effectués par la SNCF pour la formation de ses conducteurs à ce système de communication orale formalisée et je décrirai plus particulièrement l'outil sur support CD-ROM basé sur la reconnaissance vocale, créé spécifiquement pour cette formation.

1. Rappel du contexte ferroviaire

1.1. Système de communication orale formalisée

- 3 Dans le cadre des liaisons ferroviaires transmanche, une des particularités sera l'interpénétration de réseaux, c'est-à-dire la présence d'un même conducteur, français ou anglais, sur toute la durée du trajet. Pour ce dernier, il s'agit donc de se former aux spécificités ferroviaires du réseau étranger (connaissance des lignes, règles de procédures, sécurité...), mais aussi de pouvoir communiquer par radio, dans la langue du pays, avec les personnels au sol.
- 4 Il n'y aura communication entre conducteur et personnel au sol que dans des circonstances bien définies : train à l'arrêt ; pour des raisons de sécurité qui entraînent l'application d'une procédure ferroviaire spécifique.
- 5 La langue utilisée sera celle du pays dans lequel se trouve le train.

- 6 Deux implications de première importance en découlent : les conditions extraordinaires qui imposeront la mise en œuvre d'une procédure de sécurité devraient être rares ; a fortiori, le recours à la communication orale devrait donc, lui aussi, être rare.
- 7 Ce sont les conditions particulières, liées à la sécurité, la rareté des communications radio et « l'obligation » de performance de la part des conducteurs qui ont conduit au choix d'un système de communication orale formalisée.
- 8 « Formalisé » signifie ici que tout message, hormis les variables situationnelles (n° du train, nom de gare...), est pré-déterminé, que le contenu est fixé à l'avance et que le déroulement de la communication, message par message, est pré-établi (cf. annexe 1).
- 9 Le recensement des situations qui mettent en jeu la sécurité a abouti à l'élaboration de 17 formulaires de procédures, formulaires qui précisent la conduite à tenir ou les actions à entreprendre par le conducteur. Des procédures de communication radio définissent les messages radio de prise et fin de parole, de demande de répétition, de signal d'erreur... (cf. annexe 2).

1.2. Les besoins de formation

- 10 La formation des futurs conducteurs d'Eurostar s'effectue dans trois directions
- 11 a) une formation ferroviaire, du ressort naturel de la SNCF :
- 12 les conducteurs sont formés à la conduite d'un TGV, prennent connaissance des spécificités du réseau étranger et « apprennent » la ligne sur laquelle ils vont circuler. Un simulateur de conduite, entre autres, répond aux besoins de formation dans ces trois domaines ;
- 13 b) une formation linguistique, dont la conception a été confiée à CITE Langues,¹ qui répond aux besoins de communication hors trajet : les conducteurs suivent une formation qui leur permettra d'être autonomes linguistiquement à Londres durant leur période de repos et d'assurer leur prise et fin de service en gare de Waterloo. Cette formation d'une durée de 240 heures comprend également trois séjours à Londres ;
- 14 c) une formation spécifique au système de communication orale formalisée ; pour cette formation, deux outils complémentaires ont été élaborés : un outil audio « CAP Londres » (Cours pour l'Acquisition Progressive de la Langue Orale Normalisée Du Réseau EuroStar) ; un outil CD-ROM utilisant la technique de la reconnaissance vocale.
- 15 CAP Londres travaille deux aptitudes : la compréhension et l'expression orales. Il propose un ensemble d'activités variées et progressives dans leur difficulté portant sur :
- 16 - les variables : compréhension et expression des chiffres, de l'heure, de l'alphabet international, des noms propres (villes, postes d'aiguillage...).
- 17 - les mots ou groupes de mots « difficiles » : discrimination, segmentation et prononciation de mots ou groupes de mots jugés difficiles à comprendre et à maîtriser par le conducteur. Les mots difficiles sont travaillés de façon isolée puis intégrés progressivement dans leur contexte.
- 18 - les procédures : compréhension et collationnement de chaque procédure.

2. CD-ROM fondé sur la reconnaissance vocale

2.1. Objectifs

- 19 Afin de répondre aux exigences fixées par le cahier des charges SNCF, le CD-ROM devait reproduire la réalité des situations de communication telles qu'elles seront vécues par les conducteurs dans leurs interactions avec les personnels au sol ; il s'agissait donc de créer des simulations de la réalité, au sens le plus fort du terme, c'est-à-dire que toutes les situations de communication devaient être présentes et fidèles à la réalité dans leur déroulement et dans les comportements induits.
- 20 Les objectifs étaient les suivants. Tout d'abord, à travers la pratique des dialogues en simulation, viser l'acquisition de compétences linguistiques qui se traduisent par des réflexes linguistiques face à des situations ferroviaires données et génèrent des comportements ferroviaires appropriés. Ensuite, faire valider par la machine, en lieu et place du personnel au sol, la performance (compréhension et expression orales) du conducteur.
- 21 Le système de reconnaissance vocale devait donc déterminer le degré d'acceptabilité de la production du conducteur et permettre à la situation de continuer à se dérouler (c'est-à-dire, à la machine de proposer la suite de la simulation).

2.2. Partenaires

- 22 Pour la phase de travail concernant les aspects linguistiques, le partage des tâches s'est effectué entre CITE Langues et VECSYS. Vecsys, société spécialisée dans le traitement automatique de la parole, a fourni une version adaptée de la carte de reconnaissance vocale RS 1000 utilisée dans le didacticiel.
- 23 Le développement du programme informatique après intégration des données linguistiques a été effectué par CAP GEMINI/CAP SESA, l'un des leaders mondiaux du marché des services et du conseil en informatique, maître d'œuvre du projet.

2.3. Principes de reconnaissance vocale

- 24 Les premières études sérieuses en matière de reconnaissance vocale datent des années 1950². On distingue généralement deux approches : une approche analytique et une approche globale.
- 25 L'approche analytique vise à reconnaître n'importe quel message vocal à partir de la détection et l'identification des phonèmes qu'il contient. Mais la reconnaissance des phonèmes est un problème très complexe, chaque phonème pouvant recouvrir des réalisations physiques différentes selon le contexte où il se trouve, sans parler des différences très importantes que l'on peut observer entre locuteurs. Cette approche n'a pas encore donné lieu à des retombées industrielles.
- 26 L'approche globale a été plus fructueuse et les premiers systèmes commercialisés sont apparus sur le marché au début des années 70. Ne sachant reconnaître n'importe qui prononçant n'importe quoi et de n'importe quelle façon, on s'est limité dans un premier temps à un problème beaucoup plus simple : élaborer un système capable de reconnaître

un petit vocabulaire prononcé mot à mot, par un seul locuteur. Le locuteur prononce lors d'une phase d'apprentissage les différents mots du vocabulaire qu'il désire utiliser. Le système conserve en mémoire l'image acoustique de ces différents mots. Lors de la reconnaissance, il compare l'image acoustique des mots à reconnaître avec les différentes images acoustiques des mots de référence, et choisit la plus voisine. Cette technique est connue sous le nom de « reconnaissance de mots isolés monolocuteur ». Elle correspond aux contraintes les moins sévères, pour le développeur de l'application, sur les trois axes que constituent :

- . monolocuteur --> multilocuteur
- . petit vocabulaire --> vocabulaire illimité
- . mots isolés --> parole continue

- 27 Le système de reconnaissance idéal correspondant à une abscisse maximale sur ces trois axes n'existe pas encore.
- 28 Par définition, un système monolocuteur impose un apprentissage et une identification du locuteur. En revanche, un système multilocuteur ne nécessite pas de phase d'apprentissage. N'importe quel utilisateur, s'exprimant en utilisant seulement les mots du vocabulaire prévus pour l'application, sera reconnu du système. La tâche du développeur de l'application dans ce cas est plus importante : le dictionnaire des références vocales multilocuteur est construit à partir d'algorithmes opérant sur les enregistrements d'un échantillon d'environ 50 hommes et 50 femmes prononçant le vocabulaire de l'application.
- 29 La taille du vocabulaire est une caractéristique importante de tout système de reconnaissance. Plus le vocabulaire est grand, plus les possibilités de confusion sont importantes.
- 30 Certains systèmes imposent un silence entre chacun des mots (reconnaissance de mots isolés), mais l'élocution est alors peu naturelle. Un système autorisant l'enchaînement des mots est souvent préférable, mais, dans ce cas, la saisie du vocabulaire ne suffit pas. Les enchaînements des mots du vocabulaire en séquence ou en phrase doivent être fournis au système de reconnaissance sous forme d'une syntaxe déterministe.
- 31 Les domaines d'applications sont très divers : aéronautique (dialogue pilote-avion où certaines fonctions sont commandées par la voix), industriel (saisie d'information en milieu hostile), médical (saisie de comptes rendus opératoires), armement (commande de système d'armes)...

2.4. Principes retenus pour le didacticiel

- 32 Pour l'application qui nous concerne, il s'agit d'une approche globale avec reconnaissance de mots enchaînés, multilocuteur, et avec un vocabulaire limité. Un système multilocuteur s'imposait pour deux raisons : d'une part il n'était pas envisageable de passer par une phase d'initialisation avant première utilisation par chaque locuteur, d'autre part l'objectif du didacticiel étant de mesurer le degré d'acceptabilité de l'expression du conducteur, la performance de celui-ci se devait d'être comparée, non pas à sa propre référence, mais bien à un dictionnaire de références vocales.
- 33 La reconnaissance de mots enchaînés s'imposait également car respecter un silence entre chacun des mots ne correspondait pas à la réalité de la communication.

- 34 Enfin, le champ d'application de cette reconnaissance vocale n'est pas la langue anglaise dans sa globalité, mais un sous-ensemble à la fois limité en nombre (17 scenarii) et formalisé dans son expression.

2.5. Paramètres linguistiques

- 35 Il est important de préciser que la reconnaissance vocale, jusqu'à l'élaboration de ce didacticiel, n'avait été pratiquée que sur la langue maternelle du locuteur. La reconnaissance en langue seconde, ici l'anglais pour des locuteurs français, représentait donc pour la société Vecsys une première.
- 36 La liste des paramètres linguistiques à prendre en compte a été établie par CITE Langues et nous avons retenu tous les paramètres spécifiques à la langue et spécifiques au locuteur qui méritaient d'être soumis à l'appréciation de notre partenaire. Ensuite, une réflexion commune, tenant compte des exigences du cahier des charges de la SNCF et des possibilités de la carte vocale, a conduit à effectuer des choix méthodologiques et à fixer les différentes étapes de l'élaboration du système. Les principaux paramètres linguistiques qui ont eu une influence, à quelque niveau que ce soit, sur l'élaboration du système sont les suivants : les caractéristiques vocales du locuteur, l'accent, l'intonation, le débit, l'articulation, la prononciation et l'accentuation des mots.
- 37 De plus, des phénomènes tels que hésitations et reprises, oubli d'un mot ou « silence » ont eux aussi imposé des choix dans la collecte des données. Je préciserai dans la description des différentes phases de travail quels paramètres sont intervenus et comment.

3. Étapes d'élaboration

3.1. Découpage en unités

- 38 Pour répondre aux exigences techniques d'une part et aux objectifs de l'évaluation d'autre part, chaque scénario (qui comprend la procédure de communication dans laquelle s'insère le formulaire de procédure) a été découpé en unités prises en compte ultérieurement pour l'enregistrement et/ou l'évaluation.

3.1.1. Découpage en énoncés

- 39 La carte vocale utilisée dans le didacticiel peut traiter des énoncés d'une durée maximale de 25 secondes. C'est-à-dire que la reconnaissance vocale peut s'effectuer sur un énoncé de 25 secondes maximales. Dans un premier temps, il s'agissait donc de délimiter ces énoncés qui allaient correspondre à une unité d'enregistrement, c'est-à-dire que chaque énoncé va être suivi d'une pause lors de l'enregistrement, et à une unité de validation : le conducteur sera évalué après chaque énoncé lors de l'utilisation du didacticiel et son énoncé sera alors validé ou non (cf. annexe 3).
- 40 Pour les messages courts, inférieurs à 25 secondes, la délimitation de l'énoncé était aisée et systématique. En revanche, pour les messages supérieurs à 25 secondes, le découpage ne pouvait s'effectuer en fonction de la « durée » uniquement, mais devait tenir compte de critères supplémentaires. Il apparaissait essentiel de conserver des unités de sens en particulier, et de ne pas marquer de pause à des endroits qui, soit risquaient de déformer le message, soit risquaient d'entraîner une élocution peu naturelle. Pour nous approcher

au mieux de la situation réelle de communication, nous avons donc respecté les unités intonatoires et favorisé les pauses « naturelles » qui correspondent à des reprises de souffle.

3.1.2. Découpage en segments

- 41 L'énoncé, tel qu'il vient d'être défini, est l'unité maximale qui sera soumise à validation lors de l'utilisation du didacticiel. L'évaluation proprement dite portera sur des unités de reconnaissance à l'intérieur de l'énoncé, appelées segments, d'une durée maximale de 3 secondes. À partir de cette évaluation segment par segment, la machine validera ou non l'énoncé en fonction d'un facteur de pondération attribué à chaque segment, les segments sécuritaires recevant un facteur plus élevé.
- 42 Chaque énoncé devait donc être subdivisé en segments qui seraient individuellement évalués par la machine (cf. annexe 3). Ce nouveau découpage n'avait aucune répercussion sur les modalités d'enregistrement, mais servait uniquement à introduire dans la machine les unités de reconnaissance.
- 43 Lors de la situation de communication réelle, chacun des interlocuteurs aura sous les yeux le formulaire de procédure et bénéficiera donc d'un support écrit qui l'aidera à la fois dans son expression et sa compréhension. (cf. annexe 1). Le choix d'une communication orale formalisée montre ici ses avantages : chacun des interlocuteurs dispose du même document qui impose un contenu connu à la fois de l'un et de l'autre. Seules les variables restent à définir et elles ont la double caractéristique d'être à la fois hautement sécuritaires et non-prédéterminées. Elles doivent donc recevoir un traitement particulier tant dans la formation que dans l'évaluation. Par conséquent, chaque variable (ou élément de variable) a été isolée en segment et soumise à évaluation autonome à l'intérieur de l'énoncé. Les autres segments correspondent à des unités de sens d'une durée maximale de 3 secondes.

3.2. Choix de l'échantillon

- 44 Le découpage de chaque scénario en énoncés et en segments permettait d'intégrer dans la machine les différents éléments soumis à reconnaissance et évaluation. Il s'agissait alors de rassembler un échantillon de locuteurs à partir duquel serait construit le dictionnaire de références vocales. Certains paramètres linguistiques cités précédemment vont intervenir directement dans le choix de l'échantillon. Tout d'abord, le nombre de locuteurs a été fixé à 100, nombre minimal déterminé par Vecsys. Cet échantillon devant permettre d'évaluer la production des conducteurs en la comparant à des références enregistrées a été divisé en trois bases de locuteurs constituées comme suit : 20 anglophones, 33 anglicistes (enseignants d'anglais) et 47 apprenants. Les futurs conducteurs étant tous de sexe masculin, seules des voix masculines ont fait partie du dictionnaire de références vocales.
- 45 L'âge des locuteurs se situait entre 28 et 45 ans, correspondant à l'éventail des âges des conducteurs et tous les timbres de voix ont été acceptés.
- 46 Pour la base des anglophones et des anglicistes, aucune sélection particulière n'a été opérée, tout accent pouvant être représenté. Pour les apprenants, un niveau débutant n'a pas été accepté car non représentatif des futurs utilisateurs, mais hormis cette restriction, tous les niveaux étaient représentés.

- 47 Il est peut-être utile de préciser ici le fonctionnement de la reconnaissance vocale en rapport avec la constitution de l'échantillon : dans chaque base de locuteurs, l'ordinateur va effectuer pour chaque segment un moyennage et déterminer une forme de référence à laquelle sera ensuite comparée la production du conducteur ; la machine sera alors en mesure de lui indiquer si sa production est identique à ou se rapproche le plus de la base 1, 2 ou 3.

3.3. Conduite des enregistrements

- 48 Les enregistrements des 100 locuteurs se sont déroulés dans les locaux de CITE Langues qui avait aménagé un mini studio pour la circonstance. Les enregistrements étaient des enregistrements numériques avec sauvegarde sur disque optique. L'ensemble du dispositif comprenait : un ordinateur PC muni de la carte vocale, un logiciel de saisie, un lecteur enregistreur de disque optique et un combiné téléphone identique à celui utilisé pour les communications radio à bord du TGV.
- 49 La partie « conducteur » de chacun des 17 scénarii a été enregistrée par les 100 locuteurs afin de constituer le dictionnaire de références vocales. Neuf anglophones, originaires de régions différentes, ont enregistré la partie « *signalman* » des scénarii afin d'incorporer cette partie du dialogue au didacticiel pour reproduire la situation de communication.
- 50 Pour faciliter l'enregistrement, tout ce qui devait être dit s'inscrivait sur l'écran de l'ordinateur, énoncé par énoncé, en conformité avec le découpage initial en énoncés inférieurs à 25 secondes. L'unique tâche du locuteur était de parler dans le combiné, les manipulations d'instruments étant effectuées par un consultant de CITE Langues. Pour permettre des vérifications ultérieures et une analyse de l'échantillon, le nom et l'âge du locuteur ainsi que la base à laquelle il appartenait (anglophone, angliciste ou apprenant) étaient enregistrés ainsi que le nom du consultant qui effectuait l'enregistrement. Cette dernière précaution devait permettre, le cas échéant, de mettre à jour des anomalies liées éventuellement à une mauvaise manipulation des instruments par l'un ou l'autre et donc d'y remédier.
- 51 C'est dans cette phase d'enregistrement que les paramètres linguistiques vont revêtir une importance particulière. Des choix ont dû être effectués, liés d'une part à la capacité même du système de reconnaissance vocale et, d'autre part, aux objectifs du didacticiel. Par ailleurs, certains principes linguistiques validés par les trois partenaires ferroviaires et intégrés dans la charte Qualité du projet Eurostar devaient être respectés : ils stipulaient en particulier que soient respectés par chaque interlocuteur (conducteur et *signalman*) les deux principes suivants : ne pas parler plus rapidement qu'en conversation normale et prononcer chaque mot clairement.
- 52 Par conséquent, parmi les consignes données à chaque locuteur avant l'enregistrement se trouvaient des consignes concernant le débit (vitesse normale) et l'articulation (articuler chaque mot en gardant néanmoins l'enchaînement naturel des mots dans la phrase). Lors de l'élaboration du système de communication orale formalisée, il avait été spécifié que l'intonation ne devait jouer aucun rôle dans le sens des messages et donc ne pas véhiculer en tant que telle de valeur communicative. Tout en respectant ce principe, il n'était toutefois pas concevable de demander aux anglophones et aux anglicistes en particulier de « gommer » l'aspect intonatif. Il leur a donc été précisé de ne pas mettre d'emphasis particulière et de s'exprimer de façon relativement neutre tout en conservant un degré optimal d'authenticité.

- 53 Il était prévisible, car naturel, que certaines phases d'enregistrement donnent lieu à des hésitations, des reprises ou des oublis de mots et il était donc nécessaire de prévoir la conduite à tenir dans ces circonstances. Pour le système de reconnaissance vocale, les reprises et les oublis, tout comme les erreurs de mot (un mot pour un autre), déformaient la référence et par conséquent ne pouvaient être acceptés. Tout énoncé qui présentait une de ces anomalies devait donc être ré-enregistré.
- 54 En ce qui concerne les « blancs » et les « silences », un silence initial même prolongé n'avait pas d'incidence puisque l'enregistrement numérique était déclenché par le premier son émis et non par le simple appui sur la touche enregistrement. Les blancs naturels entre groupes de mots étaient acceptables, le système fonctionnant par reconnaissance de segment séparément. Les blancs prolongés ou un débit trop lent ayant pour conséquence de faire dépasser la limite des 25 secondes entraînaient systématiquement le ré-enregistrement de l'énoncé.
- 55 Les apprenants ont reçu les mêmes recommandations en ce qui concerne le débit et l'articulation que les deux autres bases de locuteurs, la consigne spécifique les concernant étant de parler « comme ils savaient ». Les erreurs de prononciation ou d'accentuation n'ont fait l'objet d'aucun traitement particulier dans cette phase d'enregistrement, l'objectif étant de constituer un échantillon de références, à la fois acceptables et non acceptables.

3.4. Notation des segments

- 56 Le découpage en segments de chaque énoncé avait pour but de délimiter des entités sur lesquelles le système de reconnaissance vocale fournirait une indication quant à la qualité de « prononciation ». Le fonctionnement de ce système qui se limite à une comparaison de chaque segment aux trois bases de références multilocuteurs permet d'indiquer si le segment prononcé s'intègre ou s'identifie à l'échantillon de telle ou telle base. C'est-à-dire qu'en l'état actuel des connaissances, le système est capable d'indiquer à l'apprenant si sa production correspond plutôt à une production d'anglophone, d'angliciste ou d'apprenant. Une évaluation plus fine était nécessaire, en particulier par rapport à la base des apprenants très hétérogène.
- 57 D'un point de vue pédagogique, il ne semblait pas judicieux de retenir une notation trop étendue (de 0 à 100 par exemple) dans la mesure où le conducteur ne serait pas forcément à même d'apprécier dans sa prestation une différence de note de 35 à 45 par exemple. Le principe d'une évaluation sur 5 niveaux a donc été retenu, de 0 à 4, niveaux assimilés à un code couleur dans le didacticiel. La détermination des niveaux s'est effectuée de la manière suivante :
- 58 - la machine a attribué systématiquement à tous les segments prononcés par un locuteur de la base des anglophones le niveau 4 (vert foncé).
- 59 - la machine a attribué systématiquement à tous les segments prononcés par un locuteur de la base des anglicistes le niveau 3 (vert clair).
- 60 - pour la base des apprenants, l'évaluation a été réalisée par CITE Langues, pour chacun des 47 locuteurs, segment par segment, pour les 17 scénarii. Nous avons donc ré-écouté les enregistrements et avons introduit directement dans l'ordinateur la notation 0, 1, 2, 3 ou 4 pour chaque segment.

- 61 La grille de notation, ré-utilisée ensuite dans le didacticiel, peut être schématisée de la façon suivante :
- 4 (vert foncé) : prononciation excellente
 - 3 (vert clair) : bonne prononciation
 - 2 (jaune) : accent français prononcé
 - 1 (orange) : erreur(s) de prononciation
 - 0 (rouge) : incompréhensible par un Anglais
- 62 Tout en laissant place à une part de subjectivité, cette grille a donné satisfaction : des notations en commun entre les différents consultants de CITE Langues, préalables à des notations effectuées individuellement, ont permis de définir des critères plus précis d'appréciation sur certains segments et d'aboutir à une bonne homogénéité de notation.
- 63 Une fois cette notation intégrée, la machine était alors en mesure d'attribuer, pour tout nouveau locuteur, une notation de 0 à 4. Toutefois, afin de vérifier le bon fonctionnement de la machine, Vecsys a procédé à l'analyse suivante : afin de simuler l'utilisation du système par un nouveau locuteur et donc d'évaluer le système, chaque locuteur a été sorti tour à tour du corpus et ré-introduit pour le comparer aux 99 restants et ainsi obtenir son évaluation. Tous les locuteurs ont donc été évalués et une notation de 0 à 4 a été attribuée à chaque segment prononcé. C'est ainsi que la fiabilité de la machine a pu être testée, en comparant la notation, segment par segment, locuteur par locuteur, de la machine et de CITE Langues.
- 64 Pour la base des anglophones et des anglicistes, notés a priori 4 ou 3, une réaffectation s'est opérée sur l'ensemble de la grille de notation. La machine a ainsi déclassé par exemple certains segments d'anglophones porteurs d'un accent très marqué, segments qui se différenciaient alors de la moyenne établie. De la même manière, certains segments d'anglicistes ont été sur-classés et notés 4 car la similitude s'établissait avec la base des anglophones et non plus avec celle des anglicistes ; d'autres segments ont été déclassés selon le même principe.
- 65 Sur l'ensemble des locuteurs et des segments, si l'on considère un écart de 1, le système de reconnaissance vocale utilisé se révèle fiable à environ 90 %, ce qui correspond à la marge moyenne d'erreur généralement trouvée et acceptée sur les systèmes existants, marge difficile à réduire pour l'instant compte tenu des possibilités techniques.

3.5. Réalisation du didacticiel

- 66 Cap Gemini s'est chargé de l'intégration de toutes les données de type son, photo, vidéo et séquences sonores et a élaboré le produit final selon les spécifications requises par la SNCF, spécifications que je ne détaillerai pas ici. L'outil de formation mis à la disposition des conducteurs est constitué des éléments suivants :
- une machine compatible PC 386 SX 20
 - un disque dur
 - une carte de restitution vidéo DVI de la firme Intel
 - un écran tactile et/ou une souris
 - un lecteur de CD-ROM pour lire le disque sur lequel se situent les données de l'application
 - une carte de reconnaissance vocale RS 1000
 - un combiné téléphonique de type ELNO
 - un dispositif de « brouillage » intégré dans le socle du combiné pour simuler les différentes qualités des liaisons radio.

4. Utilisation du didacticiel

- 67 Ce didacticiel s'insère sur un poste de travail où figurent déjà trois autres didacticiels, que je n'ai pas mentionnés car présentant des objectifs différents. À l'allumage, le menu présente le choix des didacticiels et l'appui de la touche « Radio Sol Train International » installe ce dernier.
- 68 L'entraînement à la communication radio se fait parallèlement à la formation ferroviaire : les conducteurs étudient donc les formulaires de procédures et leurs implications ferroviaires avant de venir s'entraîner avec le didacticiel.
- 69 Afin de simuler la situation réelle de communication, le conducteur dispose devant lui du formulaire de procédure qu'il suit et remplit comme il sera amené à le faire le jour de son utilisation. Le choix du numéro de la procédure lance le scénario et l'écran affiche une représentation de la platine radio utilisée à bord du TGV (cf. annexe 4). Le conducteur appuie sur la touche « SG » et l'écran affiche ensuite « parlez ». Cet ordre, dans la réalité, est initialisé par le *signalman* depuis le régulateur suite à l'appel du conducteur via la touche SG. Le dialogue s'établit alors entre le conducteur et le *signalman*.
- 70 Le conducteur, avant de parler, doit appuyer sur l'alternat du combiné, il relâchera l'alternat après chaque énoncé. Dans la réalité de la communication, l'appui de l'alternat sera maintenu durant le temps complet de chaque prise de parole, mais, dans la simulation, le découpage en énoncés de 25 secondes maximales implique l'utilisation de l'alternant pour chaque énoncé. Il est intéressant de préciser ici qu'une erreur dans la manipulation de l'alternat est signalée en tant que telle (appui prématuré ou tardif) et l'écran affiche « erreur d'alternat, recommencez ».
- 71 Après chaque énoncé, la machine donne son évaluation en inscrivant sur l'écran soit :
- 72 - « exact » : dans ce cas, le scénario continue et la voix du *signalman* poursuit le dialogue ;
ou,
- 73 - « améliorez votre prononciation » ; le conducteur a alors trois possibilités :
- il répète tout de suite la phrase ;
 - il clique sur l'icône « oreille » et il ré-entend la phrase à prononcer avant de la répéter ;
 - il clique sur l'icône ? (dans un triangle) et l'évaluation segment par segment apparaît sur l'écran.
- 74 Il a la possibilité de répéter deux fois l'énoncé sans demander l'affichage de l'évaluation segment par segment, mais, à la troisième répétition, si l'énoncé n'est toujours pas jugé « exact », l'affichage se fait systématiquement (cf. annexe 5).
- 75 Sur l'écran, les différents segments apparaissent dans la couleur qui correspond à leur évaluation et sur 5 étages, eux aussi représentatifs de l'évaluation (vert foncé en haut, rouge en bas). Le conducteur connaît alors les segments qui posent problème et peut se concentrer, lors de la répétition, sur ceux-ci.
- 76 Le code couleur est expliqué sur un écran spécifique : en cliquant sur un segment de couleur et sur l'icône ? (dans un rectangle), l'écran explicatif apparaît et indique la signification globale de la couleur ainsi que les raisons potentielles des erreurs.
- 77 Quel que soit le résultat de l'évaluation, après trois répétitions successives du même énoncé, le scénario continue. Cette option a été choisie afin de ne pas lasser le conducteur

et de ne pas provoquer de blocage psychologique. Si des constantes se manifestent dans ses erreurs, le conducteur peut se reporter à l'outil audio CAP Londres et travailler systématiquement ses points faibles avant de revenir utiliser le didacticiel.

- 78 Le dernier élément qui intervient dans la validation de l'énoncé est le facteur de pondération appliqué à chaque segment. Un facteur 1, 2 ou 3 a été attribué à chaque segment selon sa valeur sécuritaire : 3 très sécuritaire, 2 moyennement, 1 peu sécuritaire. Toutes les variables ont reçu le facteur 3 car hautement sécuritaires. Ce facteur de pondération intervient de la manière suivante : tant qu'un segment doté du facteur 3 est en zone jaune, orange ou rouge, l'énoncé sera sanctionné par l'affichage « améliorez votre prononciation », quelles que soient les couleurs des autres segments. En revanche, un énoncé où toutes les variables sont en vert pourra être considéré « exact » même si d'autres segments, pondérés 1 ou 2, sont d'une autre couleur, sauf rouge. La machine effectue automatiquement le calcul selon un modèle mathématique qui a été introduit et qui tient compte du nombre de segments dans l'énoncé et de la valeur sécuritaire de chacun. Une tolérance est donc acceptée qui favorise le déroulement du scénario.
- 79 À l'issue du travail d'un scénario, il est possible de demander une évaluation globale (cf. annexe 6). Une évaluation en cours de scénario ou chronologique est également disponible.
- 80 Le didacticiel est installé dans le centre de formation depuis un mois environ et a suscité quelques commentaires de la part des utilisateurs : l'intérêt de cet outil dans la formation est indéniable et le didacticiel est fortement apprécié. Toutefois, les conducteurs le jugent « sévère » dans son évaluation. Cette remarque appelle quelques commentaires : le didacticiel a pour objectif de former les conducteurs en vue de situations qui ne devraient se présenter que rarement et qui risquent d'être associées à des facteurs de stress. Il est donc prévisible qu'une déperdition s'effectue entre la formation et le moment d'utilisation, même si la formation est continue c'est-à-dire que l'entraînement se poursuit et est régulier. Il est par conséquent indispensable, lors de cet entraînement, de viser une qualité maximale pour que, le moment venu, une perte de qualité ne soit pas préjudiciable au dialogue et n'entrave pas son déroulement.

Conclusion

- 81 Cet outil de formation fondé sur la reconnaissance vocale est incontestablement un avantage dans la formation des futurs conducteurs du TGV Eurostar. Tout en concédant que tous les problèmes techniques ne sont pas encore résolus, nous n'en constatons pas moins que le système atteint un bon degré de fiabilité et qu'il remplit pleinement ses objectifs. Pour l'avoir utilisé, je peux témoigner du fait qu'il m'a fortement impressionnée, même si ma voix féminine, non reconnue par la machine, a été sévèrement sanctionnée !

BIBLIOGRAPHIE

- Fade, P. 1993. « Le TGV Transmanche : langue, culture et communication ferroviaires ». Communication faite au colloque européen « L'effet tunnel » organisé par l'Association Médiances à Paris, 3-5 février 1993.
- Fade, P. 1993. « Railspeak : outil de communication orale spécifique pour le TGV Transmanche ». *ASp* 1, 449-461.
- Fade, P. 1993. « Le TGV Transmanche : contraintes contextuelles et négociations ». *ASp* 2, 345-354.
- Mariani, J. 1990. « Reconnaissance automatique de la parole : progrès et tendances ». *Traitement du signal* 7/4, 239-266.
- Prouts, B. 1993. « Synthèse et reconnaissance de la parole : état de l'art ». *Techmémoires/Mémoires des commissions techniques des ingénieurs et scientifiques de France* 27.

ANNEXES

Annexe 1

FORMULAIRE DE PROCEDURE 007

Signal box :	Date :	Time :
<small>(name)</small>	<small>(dd/mm/yy)</small>	<small>(hh:mm)</small>
Instruction is given to driver of train..... <div style="text-align: right;"><small>(number)</small></div> to obey emergency speed restriction of between : <div style="text-align: right;"><small>(mph)</small></div>		
1. <input type="checkbox"/> Stations and <div style="text-align: center;"><small>(name)</small> <small>(name)</small> </div>		
2. <input type="checkbox"/> Signals and <div style="text-align: center;"><small>(number)</small> <small>(number)</small> </div>		
3. <input type="checkbox"/> m. ch. and m. ch. <div style="text-align: center;"><small>(number)</small> <small>(number)</small> <small>(number)</small> <small>(number)</small> </div>		
4. <input type="checkbox"/> and <div style="text-align: center;"><small>(name)</small> <small>(name)</small> </div>		
XX		
Instruction : <div style="text-align: center;"><small>(number)</small></div>		

Annexe 2

SCENARIO 004

DRIVER :

Victoria signals : This is train - nine, zero, zero, seven-
Standing at signal - Victor, Sierra, one, four, four-
Over.

SIGNALMAN :

Train - nine, zero, zero, seven-
This is Victoria signals.
Prepare procedure -zero, zero, four-
Over.

DRIVER :

Received.
Wait.

Ready for procedure -zero, zero, four-
Over.

SIGNALMAN :

Authority is given to driver of train - nine, zero, zero, seven-
to pass signal -Victor, Sierra, one, four, four- at danger,
and examine the line as far as signal -Victor, Sierra, one, nine, two-

Reason : track circuit failure.
And report findings.
Over.

DRIVER :

Received.
Authority is given to driver of train - nine, zero, zero, seven-
to pass signal -Victor, Sierra, one, four, four- at danger,
and examine the line as far as signal -Victor, Sierra, one, nine, two-

Reason : track circuit failure.
And report findings.
Over.

Annexe 3

Légende :
- Un double interligne sépare deux énoncés.
- /segment/

SCENARIO 004

DRIVER :

Victoria signals/ : This is train/ - nine/, zero/, zero/, seven/-
Standing at signal/ - Victor/, Sierra/, one/, four/, four/-
Over/.

SIGNALMAN :

Train - nine, zero, zero, seven-
This is Victoria signals.
Prepare procedure -zero, zero, four-
Over.

DRIVER :

Received/.
Wait/.

Ready for procedure/ -zero/, zero/, four/-
Over/.

SIGNALMAN :

Authority is given to driver of train - nine, zero, zero, seven-
to pass signal -Victor, Sierra, one, four, four- at danger,
and examine the line as far as signal -Victor, Sierra, one, nine, two-

Reason : track circuit failure.
And report findings.
Over.

DRIVER :

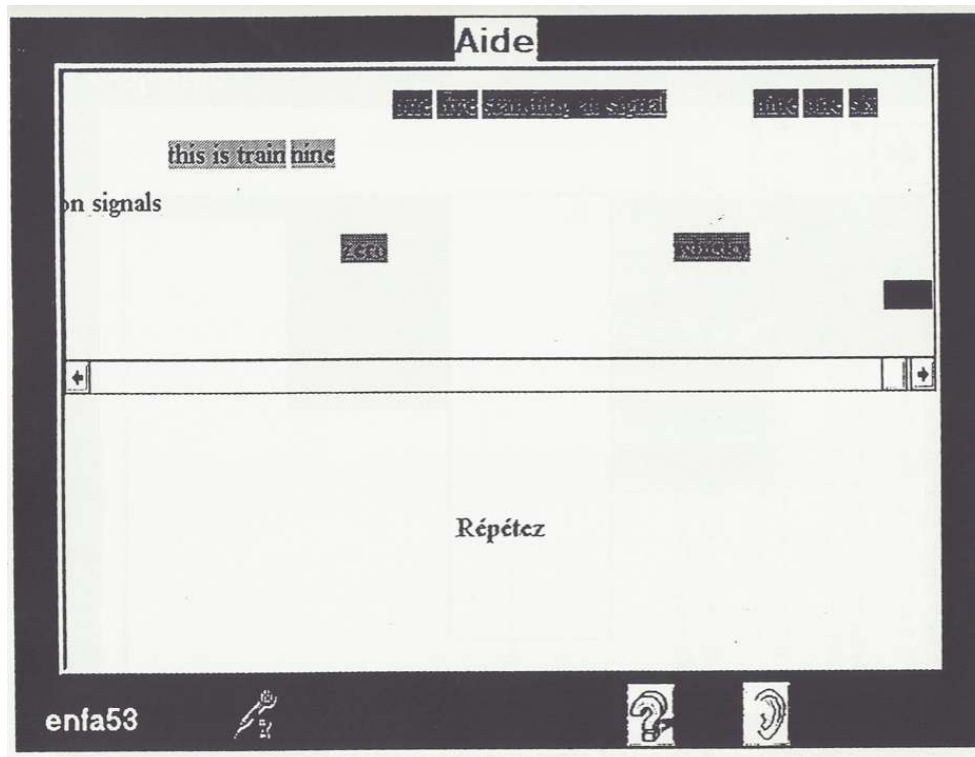
Received/.
Authority is given/ to driver of train/ - nine/, zero/, zero/, seven/- ,
to pass signal/ -Victor/, Sierra/, one/, four/, four/- at danger/ ,
and examine the line/ as far as signal/ -Victor/, Sierra/, one/, nine/, two/-

Reason/ : track circuit failure/ .
And report findings/ .
Over/.

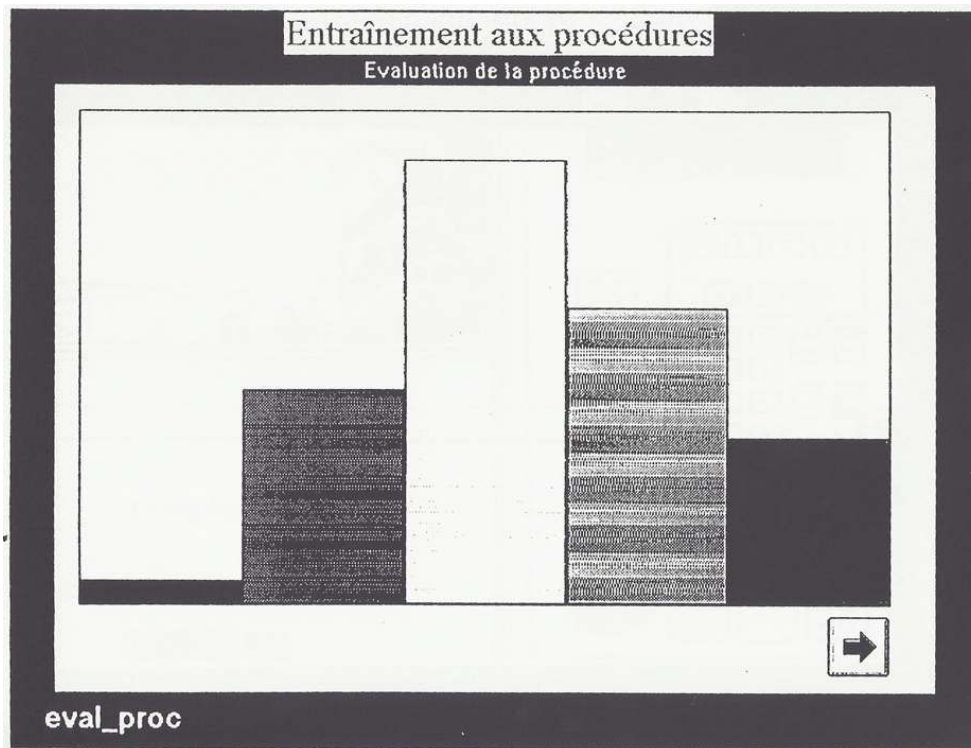
Annexe 4



Annexe 5



Annexe 6



NOTES

1. CITE Langues (Conseil Ingénierie Technologie Étude en Langues) 28, rue de Saurupt 54000 Nancy.
2. La section 2.3 est une synthèse d'un article de B. Prouts (1993).

RÉSUMÉS

Un CD-ROM utilisant la technique de la reconnaissance vocale a été élaboré dans le but de former les futurs conducteurs SNCF du TGV Transmanche au système de communication orale formalisée qui sera utilisé pour les communications radio. Les différentes étapes d'élaboration de ce CD-ROM sont décrites ainsi que le fonctionnement du didacticiel lui-même.

A CD-ROM using speech recognition was developed in order to train future SNCF Eurostar drivers to the formalized oral communication system to be used on radio transmissions. The various stages of development are described as well as the CD-ROM itself.

INDEX

Mots-clés : CD-ROM, formation, reconnaissance vocale, TGV Transmanche

Keywords : CD-ROM, Eurostar, speech recognition, training

AUTEUR

PASCALE FADE

Pascale Fade enseigne à l'Université Nancy 2 ; elle est consultante pour CITE Langues.

Pascale.Fade@univ-lorraine.fr